

Sepehr Dehdashtian

517-721-0269 | sepehr@msu.edu | [Website](#) | [Google Scholar](#) | [LinkedIn](#) | [GitHub](#)

Education

Michigan State University | Ph.D. in Computer Science Jun. 2022 – Present

- **Focus:** Responsible AI, Safety Alignment, Generative Models, Multimodal Models
- **GPA:** 4.0

Sharif University of Technology | M.Sc. in Electrical Engineering Sep. 2018 – Feb. 2021

- **GPA:** 3.87

Shahid Chamran University of Ahvaz | B.Sc. in Electrical Engineering Sep. 2014 – Aug. 2018

- **GPA:** 3.93 (ranked 1st)

Publications

FoeGlass: When Simple In-Context Learning Is Enough for Red Teaming Audio Deepfake Detectors 2026

[Sepehr Dehdashtian](#), Jacob H. Seidman, Vishnu Naresh Boddeti, Gaurav Bharaj

International Conference on Learning Representations (ICLR) 2026 (*Under Review*)

PolyJuice Makes It Real: Black-Box, Universal Red-Teaming for Synthetic Image Detectors 2025

[Sepehr Dehdashtian](#)*, Mashrur Morshed*, Jacob H. Seidman, Gaurav Bharaj, Vishnu Naresh Boddeti

Neural Information Processing Systems (NeurIPS) 2025

OASIS Uncovers: High-Quality T2I Models, Same Old Stereotypes 2025

[Sepehr Dehdashtian](#), Gautam Sreekumar, Vishnu Naresh Boddeti

International Conference on Learning Representations (ICLR) 2025 (**Spotlight: top 5%**)

Fairness and Bias Mitigation in Computer Vision: A Survey 2024

[Sepehr Dehdashtian](#)*, Ruozhen He*, Yi Li, Guha Balakrishnan, Nuno Vasconcelos, Vicente Ordonez, Vishnu Naresh Boddeti

IEEE Transaction on Pattern Analysis and Machine Intelligence (TPAMI) (*Under Review*)

The Dark Side of Dataset Scaling: Evaluating Racial Classification in Multimodal Models 2024

Abeba Birhane*, [Sepehr Dehdashtian](#)*, Vinay Prabhu, Vishnu Naresh Boddeti

ACM Conference on Fairness, Accountability, and Transparency (FAccT) 2024

Utility-Fairness Trade-Offs and How to Find Them 2024

[Sepehr Dehdashtian](#), Bashir Sadeghi, Vishnu Naresh Boddeti

IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2024

FairerCLIP: Debiasing CLIP's Zero-Shot Predictions using Functions in RKHSs 2024

[Sepehr Dehdashtian](#)*, Lan Wang*, Vishnu Naresh Boddeti

International Conference on Learning Representations (ICLR) 2024

On characterizing the trade-off in invariant representation learning 2022

Bashir Sadeghi, [Sepehr Dehdashtian](#), Vishnu Naresh Boddeti

Transactions on Machine Learning Research (TMLR) (**Featured Certification**)

Deep-Learning Based Blind Recognition of Channel Code Parameters over Candidate Sets under AWGN and Multi-Path Fading Conditions 2021
 Sepehr Dehdashtian, Matin Hashemi, Saber Salehkaleybar
 IEEE Wireless Communications Letters

Professional Experience

Research Intern at Reality Defender (Mentor: Dr. Jacob Seidman) Dec. 2024 – Present

- Developed *FoeGlass*, an algorithm to identify failure modes of audio deepfake detectors by leveraging the in-context learning capabilities of Large Language Models (LLMs).
- Developed *PolyJuice*, a red-teaming approach for image deepfake detectors that steers the text-to-image diffusion and flow-matching models to produce images capable of fooling the target classifier.

Research Assistant at Michigan State University Jun. 2022 – Present

- Developed Responsible AI algorithms to make computer vision, multimodal, and generative models fair and debiased.
- Published papers in top computer vision and machine learning conferences: NeurIPS'25, ICLR'25, ICLR'24, CVPR'24, FAccT'24.

Awards & Honors

- ICLR 2025 Spotlight Paper (Top 5%) 2025
- Interdisciplinary Inquiry and Teaching Fellowship 2025
- STEAMpower Fellowship 2024
- TMLR Outstanding Paper Award Runner-Up and TMLR Featured Certification Award 2023
- Ranked 2nd GPA the graduating class of 2021 | Sharif University of Technology 2021
- Ranked 1st GPA the graduating class of 2018 | Shahid Chamran University of Ahvaz 2018

Technical Skills

Languages: Python, C++, CUDA, Verilog, VHDL

ML Frameworks: PyTorch, PyTorch-Lightning, TensorFlow, Keras

Others: RevealJS, Git, MATLAB, FPGA

Other Projects

- Mitigating Political Bias in Pre-Trained Large Language Models (LLMs) 2023
- Visually Explaining Fair Representation Learning—A Model Perspective 2023
- Video Synopsis using OpenCV in Python 2019

Services & Activities

- Reviewer: ICLR, NeurIPS, TPAMI Jan. 2024 – present
- Mentored Student: Yilin Zheng (Master Student at MSU) Jan. 2024 – present
- Scientific-Students Associations of Clean Energy | Deputy Secretary 2016
- Scientific-Students Associations of EE Department | Member 2015